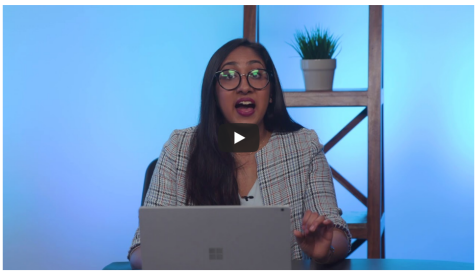
The Computer Science Perspective



Computer science terminology

As we discussed earlier, one of the simplest ways we can organize data for machine learning is in a table, like the table of clothing products we looked at earlier in this lesson:

SKU	Make	Color	Quantity	Price
908721	Guess	Blue	789	45.33
456552	Tillys	Red	244	22.91
789921	A&F	Green	387	25.92
872266	Guess	Blue	154	17.56

What are some of the terms we can use to describe this data?

For the **rows** in the table, we might call each row an **entity** or an **observation** about an entity. In our example above, each *entity* is simply a product, and when we speak of an *observation*, we are simply referring to the data collected about a given product. You'll also sometimes see a row of data referred to as an **instance**, in the sense that a row may be considered a single example (or instance) of data.

For the **columns** in the table, we might refer to each column as a **feature** or **attribute** which describes the property of an entity. In the above example, `color` and `quantity` are *features* (or *attributes*) of the products.

Input and output

Remember that in a typical case of machine learning, you have some kind of *input* which you feed into the machine learning algorithm, and the algorithm produces some *output*. In most cases, there are multiple pieces of data being used as input. For example, we can think of a single row from the above table as a *vector* of data points:

```
(908721, Guess, Blue, 789, 45.33)
```

Again, in computer science terminology, each element of the input vector (such as `Guess` or `Blue`) is referred to as an *attribute* or *feature*. Thus, we might feed these *input features* into our machine learning program and the program would then generate some kind of desired output (such as a prediction about how well the product will sell). This can be represented as:

```
Output = Program(Input Features)
```

An important step in preparing your data for machine learning is *extracting* the relevant features from the raw data. (The topic of *feature extraction* is an important one that we'll dive into in greater detail in a later lesson.)

QUESTION 1 OF 2

Have a look at this data:

ID	Name	Species	Age
1	Jake	Cat	3
2	Bailey	Dog	7
3	Jenna	Dog	4
4	Marco	Cat	12

Which of the following terms might we use to refer to the part of the table that is highlighted?

(Select all that apply.)

- [x] A *row*
- [ ] An *attribute*
- [x] An *entity*
- [x] An *instance*
- [x] An *input vector*
- [ ] A *feature*

SUBMIT

QUESTION 2 OF 2

And how about now?

ID	Name	Species	Age
1	Jake	Cat	3
2	Bailey	Dog	7
3	Jenna	Dog	4
4	Marco	Cat	12

Which of the following terms might we use to refer to the part of the table that is highlighted?

(Select all that apply.)

- [x] A *column*
- [x] An *attribute*
- [ ] An *entity*
- [ ] An *instance*
- [x] A *feature*

SUBMIT

NEXT